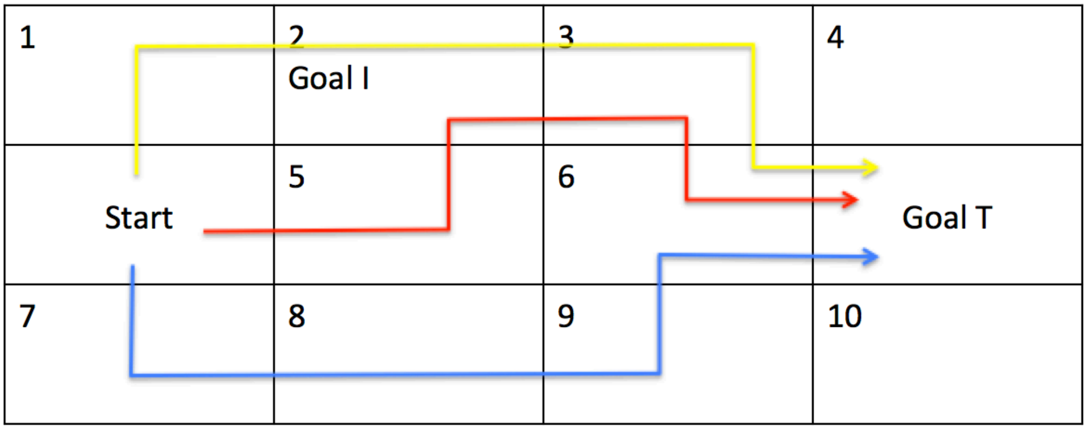
Reinforcement Learning

Suppose a reinforcement learning system is using Q-learning to learn how to navigate in an environment from a given start state. The environment has a *terminal* goal state (Goal T) that gives reward R(T)=10 and a *non-terminal* intermediate goal state (Goal I) that gives reward R(I) = 2.



Q Table:

State	Up	Down	Left	Right
1	0.0	0.0	0.0	**0.2**
2	0.0	0.0	0.0	**0.0025**
3	0.0	**0.14**	0.0	0.0
4	0.0	0.0	0.0	0.0
5	**0.2**	0.0	0.0	0.0
6	0.0	0.0	0.0	**2.71**
7	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0

The discount factor is 0.5. The learning rate is 0.1. The agent does not have uncertain actions.

The first trial is marked blue.

1. During the first trial, compute the Q-table row for state 6.

   $Q(s_6,a_{right})=0+0.1(10-0)$ = **1**

The second trial is marked red.

2. During the second trial, compute the Q-table row for state 5.

   $Q(s_5,a_{up})=0+0.1(2+0.5(0)-0)$ = **0.2**

3. During the second trial, compute the Q-table row for state 3.

   $Q(s_3,a_{down})=0+0.1(0+0.5(1)-0)$ = **0.05**

4. During the second trial, compute the Q-table row for state 6.

   $Q(s_6,a_{right})=1+0.1(10-1)$ = **1.9**

The third trial is marked yellow.

5. During the third trial, compute the Q-table row for state 1.

   $Q(s_1,a_{right})=0+0.1(2+0.5(0)-0)$ = **0.2**

6. During the third trial, compute the Q-table row for state 2.

   $Q(s_2,a_{right})=0+0.1(0+0.5(0.05)-0)$ = **0.0025**

7. During the third trial, compute the Q-table row for state 3.

   $Q(s_3,a_{down})=0.05+0.1(0+0.5(1.9)-0.05)$ = **0.14**

8. During the third trial, compute the Q-table row for state 6.

   $Q(s_6,a_{right})=1.9+0.1(10-1.9)$ = **2.71**